Introduction to Bioinformatics

Biological Networks

Department of Computing
Imperial College London

Spring 2010

Lecturer: Nataša Pržulj
natasha@imperial.ac.uk

Introduction to Cellular Networks
Intra-cellular networks
(continued)

Protein-Protein Interaction Networks

Methods for their detection (continued)

We have seen in last class:

1. Co-immunoprecipitation (CoIP, “pull-down”)
2. Yeast-2-hybrid (Y2H)
   - Data download: from Marc Vidal’s web page, Harvard Medical School: http://ccsb.dfci.harvard.edu/web/www/ccsb/
     also see databases below

Today:

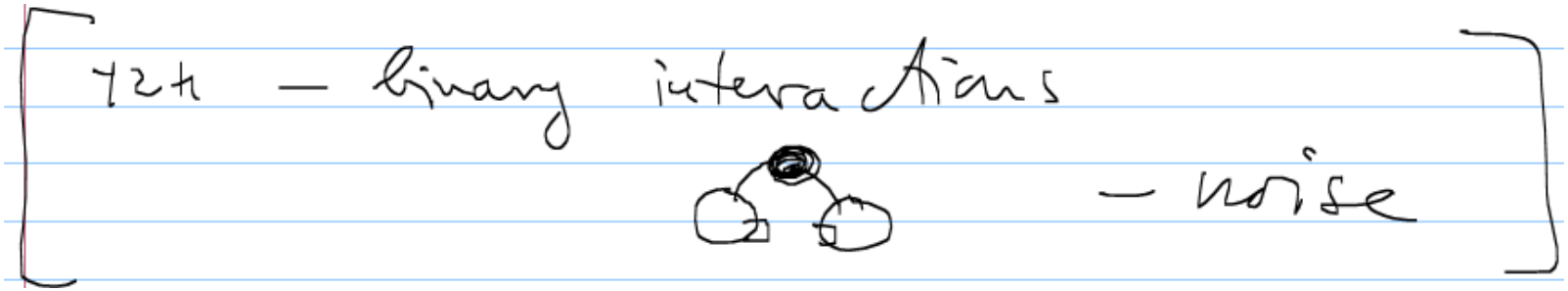
3. Mass spectrometry of purified complexes

Protein-Protein Interaction Networks Methods for their detection (continued)

3. Mass spectrometry of purified complexes

- Tag individual proteins – used as hooks to biochemically purify whole protein complexes

Asside – from last class – another source of noise in Y2H:

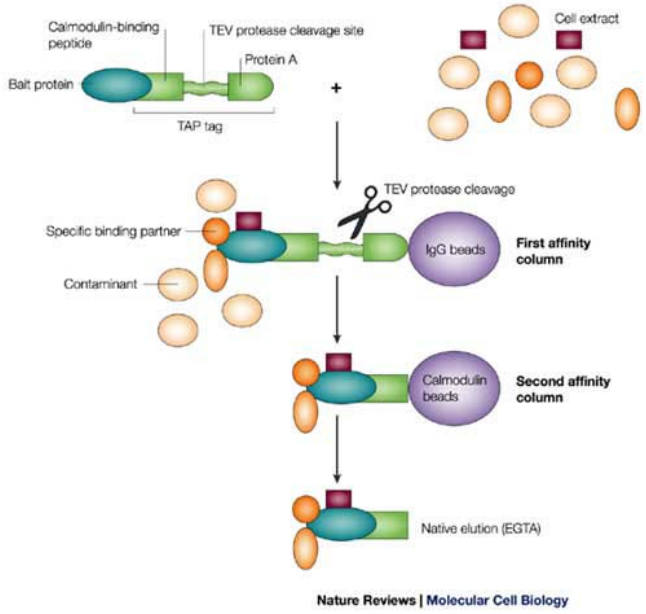


- Complexes seperated & components identified by *mass spectrometry (MS):*
  - mass spectrometry measures mass-to-charge ratio of ions

Mass spectrometry of purified complexes

There exist two main protocols:

1. *Tandem affinity purification (TAP)*
   - a fusion protein created with the ”TAP tag” that binds to beads



Mass spectrometry of purified complexes

2. *Hight-Throughput MS Protein complex (HMS-PCI)*
   - See Ho *et al., Nature* 415, 2002 (optional additional reading)
   - They used about 300 baits and about 3,600 preys in yeast

We know what proteins are in the comples, but not how they are connected

letters to nature

Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry

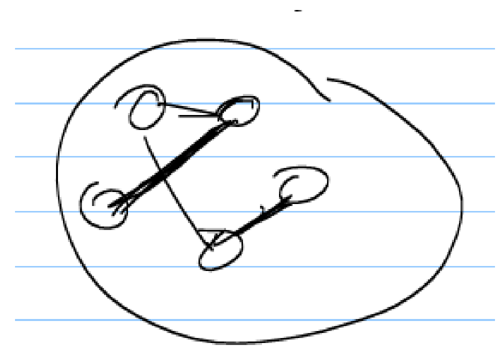
Yuen Ho*, Albrecht Gruhler*, Adrian Heilbut*, Gary D. Bader†‡, Lynda Moore*, Sally-Lin Adams*, Anna Millar*, Paul Taylor*, Keiryn Bennett*, Kelly Boutilier*, Lingyun Yang*, Cheryl Wolting*, Ian Donaldson*, Søren Schandorff*, Juanita Shewnarane*, Mai Vo*†, Joanne Taggart*†, Marilyn Goudreault*†, Brenda Muskat*, Cris Alfarano*, Danielle Dewar†, Zhen Lin†, Katerina Michalickova†‡, Andrew R. Willems†§, Holly Sassi†, Peter A. Nielsen*, Karina J. Rasmussen*, Jens R. Andersen*, Lene E. Johansen*, Lykke H. Hansen*, Hans Jespersen*, Alexandre Podtelejnikov*, Eva Nielsen*, Janne Crawford*, Vibeke Poulsen*, Birgitte D. Sørensen*, Jesper Matthiesen*, Ronald C. Hendrickson*, Frank Gleeson*, Tony Pawson†§, Michael F. Moran*, Daniel Durocher†§, Matthias Mann*, Christopher W. V. Hogue†‡, Daniel Figeys* & Mike Tyers†§

* MDS Proteomics, 251 Attwell Drive, Toronto, Canada M9W 7H4, and Staermosegaardsvej 6, DK-5230 Odense M, Denmark
† Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Canada, M5G 1X5
‡ Department of Biochemistry, University of Toronto, 1 Kings College Circle, Toronto, Canada M5S 1A8
§ Department of Medical Genetics and Microbiology, University of Toronto, 1 Kings College Circle, Toronto, Canada M5S 1A8

The recent abundance of genome sequence data has brought an

cannot be resolved by peptide-mass-fingerprinting alone, we used tandem mass spectrometry (MS/MS) fragmentation to identify unambiguously proteins in each gel slice[6]. A total of 15,683 gel slices were processed, yielding approximately 940,000 MS/MS spectra that matched sequences in the protein sequence database. Over 35,000 protein identifications were made, corresponding to 8,118 potential interactions with a set of 600 bait proteins that were expressed at detectable levels (Supplementary Information Table 1). Ubiquitous, nonspecifically binding proteins, defined empirically on the basis of frequency of occurrence (see Supplementary Information), were subtracted from the raw data set to yield 3,617 interactions with 493 baits—for further discussion of filtering criteria and mass spectrometry methodology, see Supplementary Information. This filtered data set contained 1,578 different interacting proteins representing 25% of the yeast proteome (Supplementary Information Table 2). In a preliminary direct validation of the HMS-PCI data set, 64 out of 86 interactions (74%) in a random set of new associations detected by HMS-PCI were recapitulated in immunoprecipitation-immunoblot experiments (data not shown). The HMS-PCI method was able to identify known complexes from a variety of subcellular compartments, including the cytoplasm, cytoskeleton, nucleus, nucleolus, plasma membrane, mitochondrion and vacuole (see Supplementary Information). Of all the proteins identified, 531 corresponded to hypothetical uncharacterized proteins predicted from the yeast genome sequence (Supplementary Information Table 3).

To begin to assess cellular signalling events on a proteome-wide level, we used most of the protein kinases and phosphatases encoded in the yeast genome to capture associated components. As an example, HMS-PCI analysis of the mitogen-activated protein kinase (MAPK) Kss1 identified many known components of the mating/filamentous growth pathway[7], including Ste11, Ste7, and



   - Mike Tyers lab produced the data

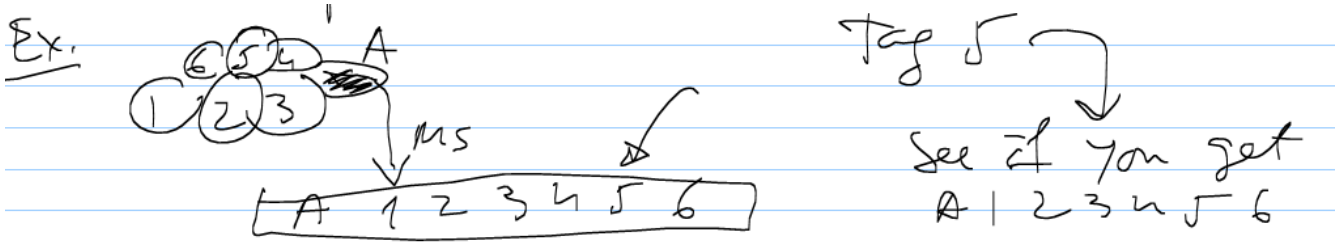
- While Y2H detects binary interactions, MS techniques detect entire protein complexes

Mass spectrometry of purified complexes

Pros and Cons of MS-based techniques for PPI detection

Good:

1. detect real complexes in their physiological settings
2. consistency check is possible by tagging several members of a complex



3. good for screening permament / stable interactions

Mass spectrometry of purified complexes

Pros and Cons of MS-based techniques for PPI detection

Bad:

1. Might miss some complexes that aren't present under given cellular conditions
2. Tagging may:
   - disturb complex formation
   - affect protein expressen levels
3. Losely associted components can be washed off during purification

Optional additional reading:
Chapter 3 of "Knowledge Discovery in Proteomics" by Wiggle and Jurisica

Other biochemical methods for PPI detection

1. LUMIER (Luminesence-based Mammalian IntERactome Mapping), by M. Barrios-Rodiles, *Science* 307, 2005
   - Maps dynamic PPIs in TGF-beta
   - ~ 15 baits used to detect ~ 300 preys

2. Correlated m-RNA (messenger RNA) expression (synexpression)
   - *m-RNA* definition: it is transcribed from DNA, carries coding information to the ribosomes where proteins are synthesized

Genetic Interaction Networks

This is a different type of network than PPI networks.

- In vivo
- Detecting interactions by ovserving the phenotypic results of gene mutations
- Example: synthetic lethality
    - *Single mutants* (mutations of single genes in an organism) yield no differences in *phenotype* (observable characteristic of an organism)
    - Mutations of **two genes** (double mutants) make cell sick or dead
    - Nodes are genes
    - Edges are drawn between pairs of genes which when mutated together make the cell sick or dead
    - Thus, these networks model functionally associated genes
    - How are these genes associated?
        - Via a PPI?
        - Over-engineering, i.e., alternative pathways for robustness to perturbations?
        - In a protein complex one mutated protein can be ”tolerated” (complex is still functional) but not two mutated proteins?

Genetic Interaction Networks

Example:

Tong *et al*. papers published genetic interaction networks, University of Toronto

In silico predictions of PPIs

- Using computational approaches to predict interactions
- Screening whole genomes for types of interaction evidence such as:
  - gene fusion (if two genes are present in one species and fused in another)
  - gene neighbourhoods (transcribed in the same time)
  - ”gene co-occurences” = ”similar phylogenetic profiles:”
    - co-occurence or absence of pair of non-homologous genes accross genomes (*homologous* genes have shared ancestry)
    - this is because if they co-occur then they might be interacting
  - Structural (in protein 3-dimensional structure) & sequence (in protein sequence) motifs (patterns repeated at frequencies higher than expected at random) within protein-protein interfaces of *known* interactions

    → By examining known interactions in this way the goal is to construct general ”rules” for protein interaction interfaces

PPI Databases

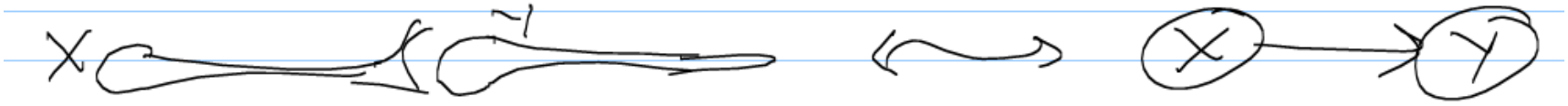
- MIPS = Munich Information Center for Protein Sequences
- YPD = Yeast Proteomics Database
- DIP = Database of Interacting Proteins
- HPRD = Human Reference Protein Database
- GRID = General Repository for Interaction Datasets
- MINT = Molecular INTeraction Database
- VirusMINT
- OHPID = Online Predicted Human Interaction Database (now called I2D = Interologous Interaction Database)

You can download PPI networks for different species from these databases.

Other Biological Networks – non cellular

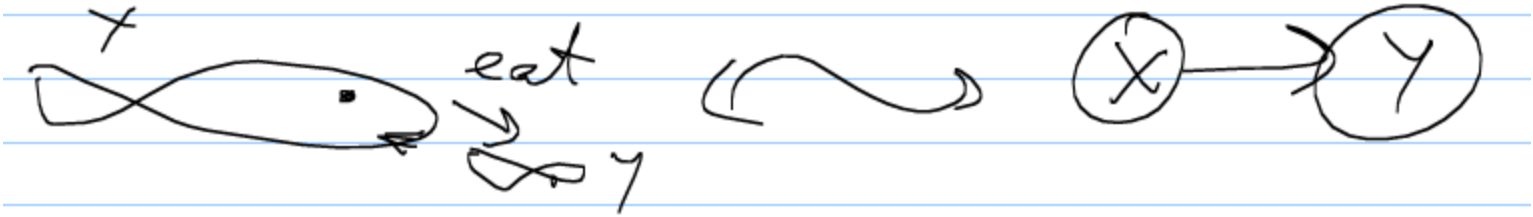
- Neuronal synaptic connection networks

  Example: neurons X and Y



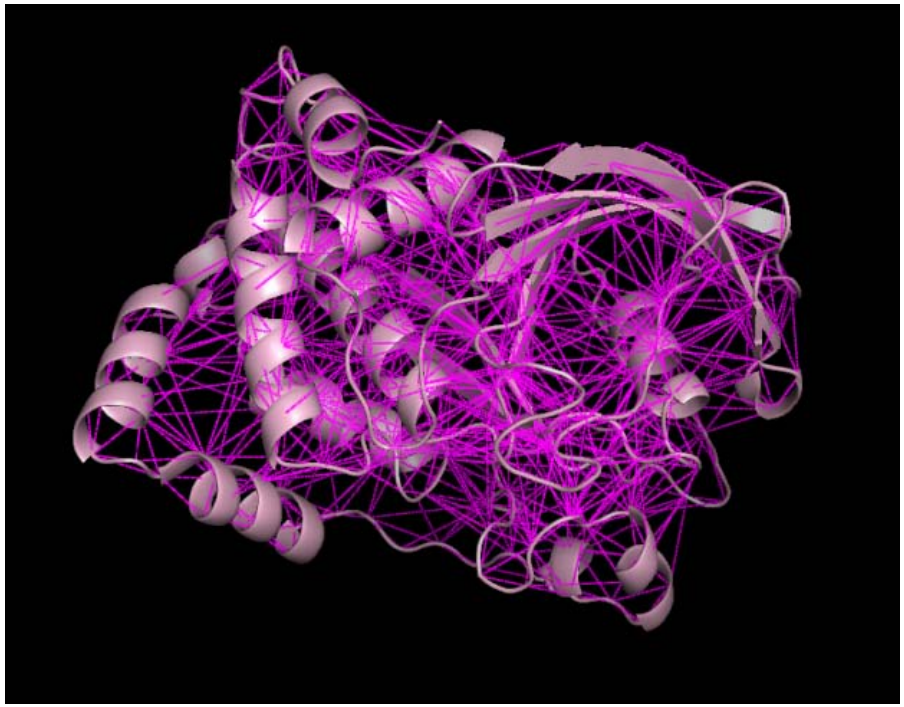
- Brain functional networks
  - nodes are brain regions
  - edges are correlations between regions that are active simultaneously during a performance of a task by the subject

- Ecological food webs

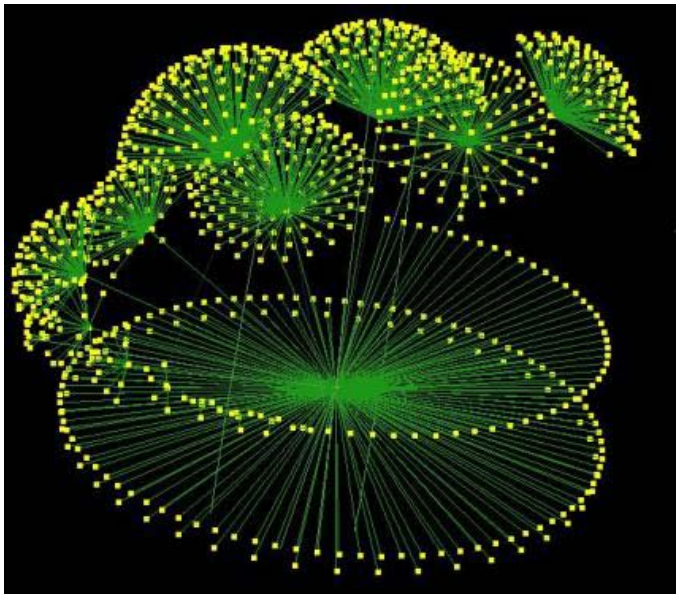


Other Biological Networks – non cellular

- Residue Interaction Graphs (RIGs)
  - they model protein structure
  - nodes are amino acid residues
  - edges are drawn between amino acids that are in close proximity in the protein's 3-dimensional structure:
    e.g., within 5 *Angstroms* (1 Å = 10^{-10} Meter)



Other Real-World Networks



- Techonlogical networks:
  - www,
  - internet,
  - electric circuits,
  - software call graphs
- Transportation networks:
  - Roads, airlines, railways
- Social networks:
  - collaborations between scientists / movie stars,
  - spread of infections disease,
  - economic networks,
  - relationships between organizations (companies, NGOs, etc.)
  - city / country trading relationships,
  - migrations,
  - disaster response networks

Other Real-World Networks

- All use similar modeling tools, BUT
- we need to be application specific
- this is because some problems might be computationally hard in general, but easy for a particular application
- E.g., finding isomorphism between *trees* (graphs with no cycles) can be done in linear time, but it is hard on graphs in general
- This is one of the reasons why it is important to find a *network model* (will be defined in next class) to which a real-world network belongs